STAT 206
6 Feb 20

$$\begin{bmatrix}0\\ \vdots\\ 0\\ \hline 1\\ \vdots\\ 1\end{bmatrix}\left.\begin{array}{c}\uparrow\\ \beta\\ \downarrow\\ \uparrow\\ \alpha\\ \downarrow\end{array}\right\}$$

prior sample size $n_0 = (\alpha+\beta)$

prior info

mean $\frac{\alpha}{\alpha+\beta} = E(\theta \mid \text{Beta}(\alpha,\beta)\ \mathcal{B})$

$$\begin{array}{c}\uparrow\\ n\\ \downarrow\end{array}\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix} = \begin{bmatrix}0\\ \vdots\\ 0\\ \hline 1\\ \vdots\\ 1\end{bmatrix}\begin{array}{c}\uparrow\\ n-s\\ \downarrow\\ \uparrow\\ s\\ \downarrow\end{array}$$

data info.

data sample size n

mean $\bar{y} = \frac{s}{n} = \frac{s}{s+(n-s)}$

$$\frac{\alpha+s}{\alpha+\cancel{s}+(\beta+n-\cancel{s})} = \frac{\overset{(\alpha+\beta)}{(w)}\left(\frac{\alpha}{\alpha+\beta}\right)+\overset{\boxed{n}}{(1-w)}\frac{s}{n}}{\alpha+\beta+n}$$

$0 \le w \le 1$

posterior mean

prior (mean)

data mean

$$(\theta \mid \mathcal{B}) \overset{\text{low information}}{\sim} \boxed{LI}(\theta)$$

$$(Y_i \mid \theta \, \overset{\text{exch}}{\wedge} \, \mathcal{B}) \overset{IID}{\sim} \text{Bernoulli}(\theta)$$

$$(i = 1, \ldots, \boxed{n})$$

---

if context implies not much known about θ external to $\mathcal{Y}$

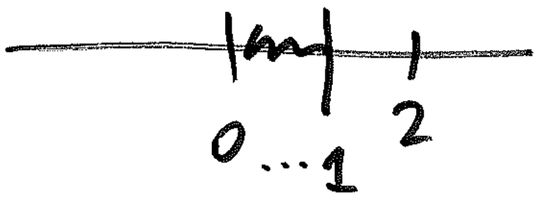
$$\text{Beta}(\overset{\theta \,|}{\wedge}\alpha, \beta) = LI(\theta) \quad \begin{pmatrix} \alpha > 0 \\ \beta > 0 \end{pmatrix}$$

$$(n_0 = \alpha + \beta)$$

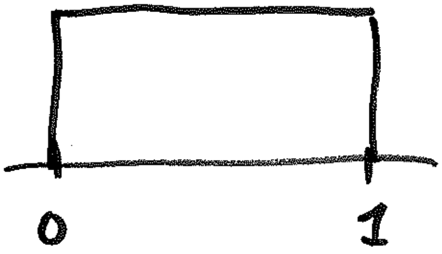
$\updownarrow$

$(\alpha + \beta)$ close to 0

$\downarrow$



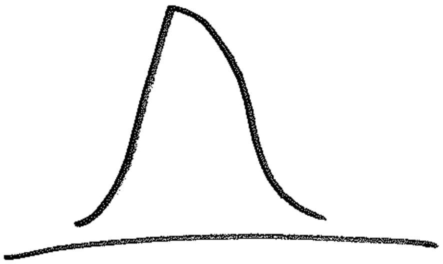
Beta(1,1) = uniform(0,1)



$p(\theta \mid \text{Beta}(\alpha, \beta)\ \mathcal{B})$

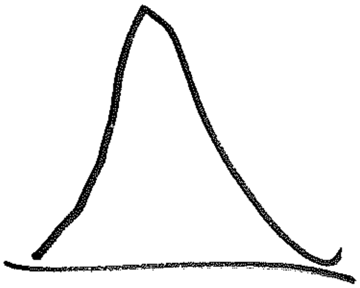
$LI(\theta \mid$ prior

prior sample size = 2

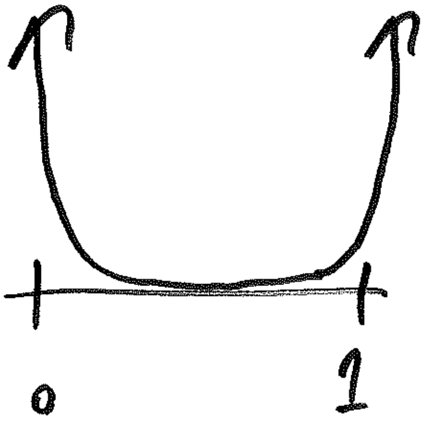


$\ell(\theta \mid \underset{\sim}{y}\ \text{exch}\ \mathcal{B})$ lik.

data sample size $n = \boxed{403}$

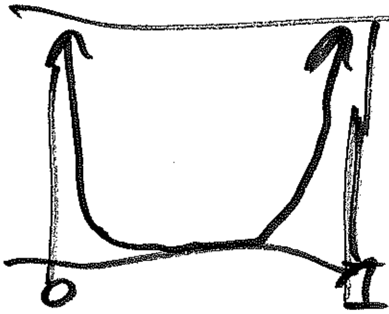


$p(\theta \mid \underset{\sim}{y}\ \boxed{\text{exch}}\ \mathcal{B}\ \text{Beta}(\alpha, \beta))$ post.



Beta$(\frac{1}{2}, \frac{1}{2})$ prior sample size 1

(1950)
(Jeffreys prior)



Beta(0, 0) ——— = 0

(Haldane prior) (1935)

$$\int_0^1 \text{Beta}(\underset{\uparrow}{0}, 0)\, d\theta = +\infty$$

such a dist. is called <u>improper</u>

& Beta(0, 0) prior with IID Bernoulli lik. f'n is an example of the use of an <u>improper prior</u>

---

newbie Bayesian advice: don't use improper priors

$\boxed{\text{p.n.s.s.} = 2\epsilon}$ for $\epsilon > 0$ & near 0

V. low IC, almost improper: Beta(ϵ, ϵ)

SD 0.0184

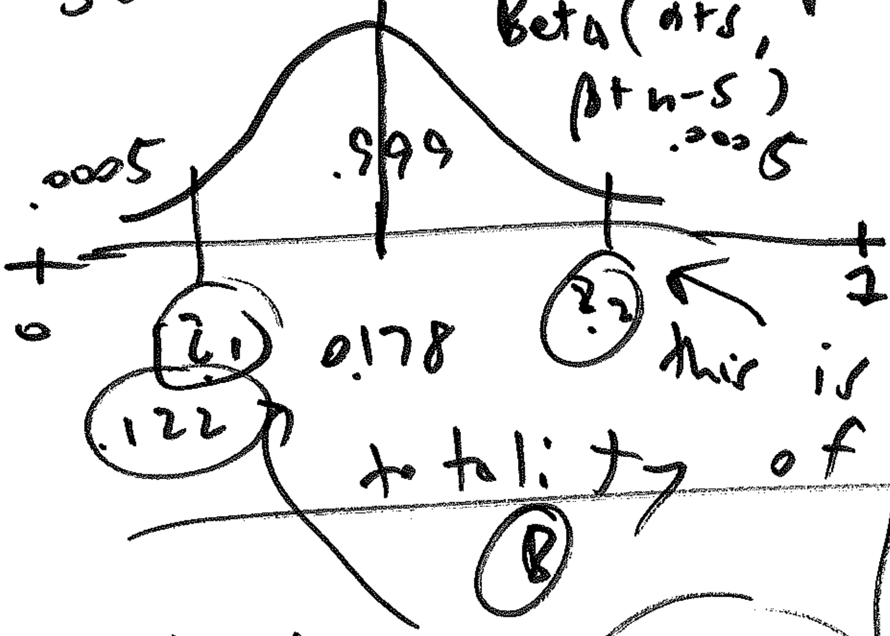
post. dist

$Beta(\alpha+s, \beta+n-s)$

$p(\theta \mid y) \text{ (exch) } Beta(\alpha, \beta)$

α ← 4.5

β ← 25.5



this is the totality of Bayesian inference

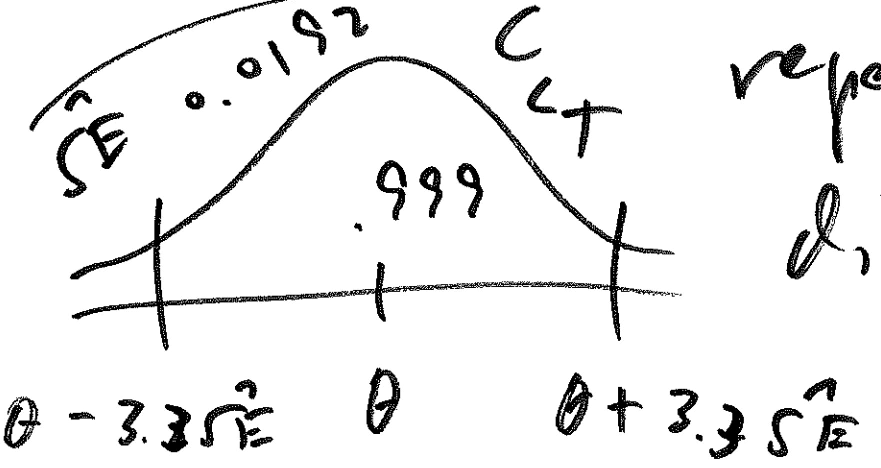
Ⓑ	Ⓕ
post. mean = 0.178	data mean ~~0.180~~ 0.179
post. SD 0.0184 $n_{post} = \underbrace{(\alpha+\beta)+n}^{433}$	$\hat{SE}(\bar{y}) =$ 0.0192 $n = 403$
99.9% credible interval (0.122, 0.242)	99.9% CI (0.116, 0.241)

$6.999 = P_B \left(?_1 \leq \theta \leq ?_2 \right)$

$\text{qbeta}(\underset{\overset{\wedge}{.9995}}{\overset{.0005}{,}}\ 4.5 + 72,\ 25.5 + 403 - 72)$ $\textcircled{?_1}$

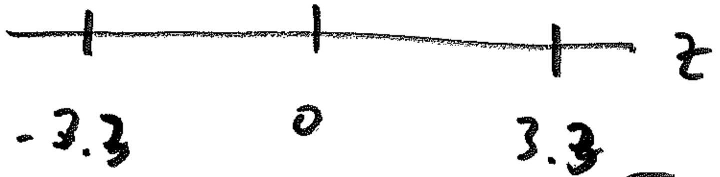
$\textcircled{?_2}$

.122



repeated-sampling dist of $\hat{\theta}$

(n = 403)



← qnorm(.9995)

99.9% CI for θ : $\hat{\theta} \pm 3.3\,\hat{SE}(\hat{\theta})$

Ⓕ

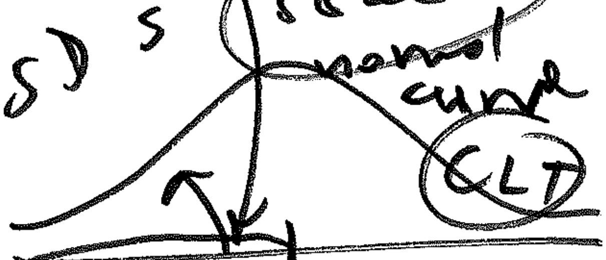
$$\boxed{c_1 e^{-c_2(\hat{\theta}_{MLE}-\theta)^2}}$$ θ (fixed)

repeated-sampling dist of $\hat{\theta}_{MLE}$ (RS) (R.V.), n large

normal curve for $\hat{\theta}_{MLE}$ in RS, centered at θ

Ⓑ

same! normal curve



$$\boxed{c_1 e^{-c_2(\theta-\hat{\theta}_{MLE})^2}}$$ $\hat{\theta}$ (fixed)

post. dist. for R.V. → θ given y & LI prior, n large

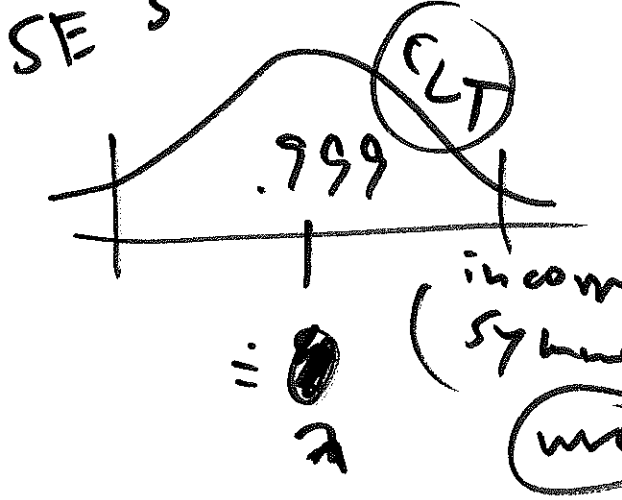
normal curve for θ, centered at $\hat{\theta}_{MLE}$

Laplace 1800

B-vM

Bernstein-von Mises thm: with large n & LI prior, frequentist inf. ≐ Bayesian inf.

Ⓕ (this week's disc. sec.)



$\hat{\lambda}_{MLE}$

RS dist of MLE, n small

simple versions of Neyman & Fisher pretend CLT always holds, even for small n

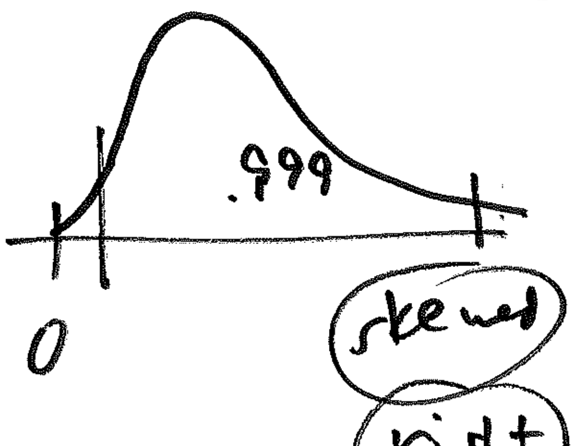
---

Ⓑ

post dist for λ given $\hat{\lambda}$ (n small)

14

& LI prior



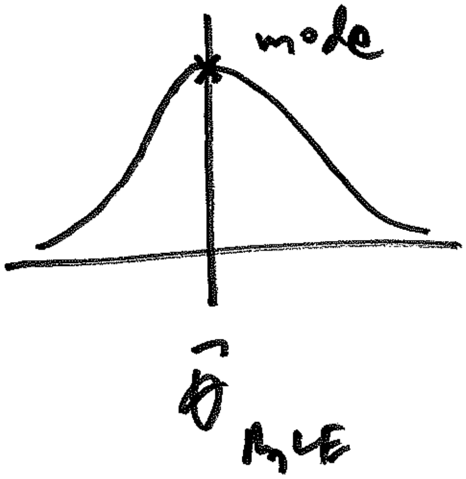
MLE story easier than Bayes

differentiation (easier)

integration (harder)

---

when B-vM holds, can therefore use freq. (MLE) as good approx to Bayes



$\ell(\theta \mid y \text{ exch. } \mathcal{B})$ B-vM

$\doteq p(\theta \mid y \text{ exch. } \mathcal{B})$

$\doteq$ normal

n large

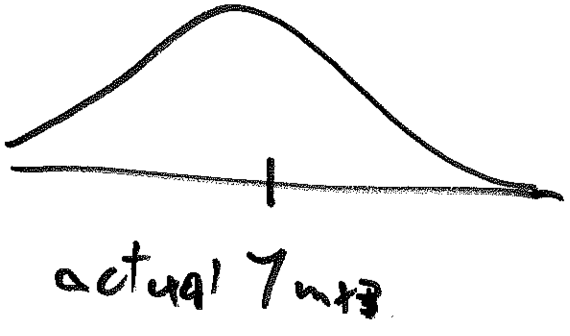
$\doteq$ post. median

$\doteq$ post. mean ← Bayesian point est.

	y	excl. stat.	ML
Ⓜ	$y_1 \ldots y_m$	modeling	(train)
Ⓥ	$y_{m+1} \ldots y_n$	validation	(test)

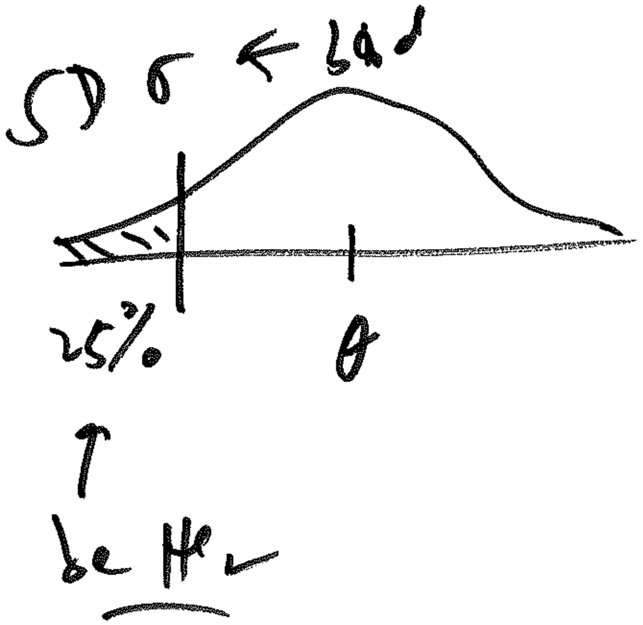
build model on Ⓜ data, check its goodness by forcing it to predict Ⓥ data :

2-way cross-validation



$p(\underline{\underline{y_{m+3}}} \mid y_1 - y_m \text{ model})$

how compare a data value with a pred. $\frac{\text{dist}}{\text{for it?}}$


SD σ ← bad
25%
θ
↑
better